Cute Little Puppies and *Nice Cold* Beers: An Information Theoretic Analysis of Prenominal Adjectives

Melody Dye (meldye@indiana.edu)[1], Petar Milin (p.milin@sheffield.ac.uk)[2], Richard Futrell (futrell@mit.edu)[3], & Michael Ramscar (michael.ramscar@uni-tuebingen.de)[4]

[1] Cognitive Science Program, Indiana University, Bloomington [2] Department of Journalism, University of Sheffield
[3] Department of Brain and Cognitive Sciences, MIT [4] Department of Linguistics, University of Tübingen

Abstract

A central goal of typological research is to characterize linguistic features in terms of both their functional role and their fit to social and cognitive systems. One longstanding puzzle concerns why certain languages employ grammatical gender. In an information theoretic analysis of German noun classification, Dye et al. (2017) enumerated a number of important processing advantages gender confers. Yet this raises a further puzzle: If gender systems are so beneficial to processing, what does this mean for languages that make do without them? Here, we compare the communicative function of gender marking in German (a deterministic system) to that of prenominal adjectives in English (a probabilistic one), finding that despite their differences, both systems act to efficiently smooth information over discourse, making nouns more equally predictable in context. We examine why evolutionary pressures may favor one system over another, and discuss the implications for compositional accounts of meaning and Gricean principles of communication.

Keywords: prenominal adjectives; grammatical gender; language comprehension; language production; language evolution; information theory; typology; formal semantics; Gricean conversational maxims

Introduction

Linguistic typologists work to define similarity and difference across languages, in an effort to establish what invariant 'universal' properties might underpin the fundamental human capacity for language, amidst remarkable diversity (Evans & Levinson, 2009). This enterprise is complicated by the fact that language is a hybrid system, which is product both of a common biological endowment (shared across languages and peoples) and of a particular ecological niche (specific to a given language).

As languages evolve, they adopt communicative strategies in response to both social and cognitive pressures, strategies which are then refined over generations of cultural transmission (Becker et al., 2009; Tomasello, 2003; Boyd & Richerson, 2005; Atkinson & Gray, 2005). In seeking to understand the limits on variation, a typologist has the unenviable task of disentangling biological imperative (Christiansen & Chater, 2008) from cultural and historical contingencies (Lupyan & Dale, 2010), such as migrations or language contact.

Information theoretic approaches to language can help clarify this problem, by setting a goalpost that is explicitly functional, rather than biological or cultural (Ramscar & Baayen, 2013). On this view, language is a communication system like any other, with the same fundamental purpose of transmitting information. A language's structural features should thus be subject to the selection pressures that govern the design of efficient digital codes (Dahl, 2004). On this read, variation among languages is the result of selective adaptations to variable circumstances, with communicative efficiency the key measure of fitness.

This is not to imply that the solutions that different languages converge on are 'equally' optimal to some pre-specified degree. Evolutionary processes achieve local—rather than global—optima, and are chained to their particular historical lineage (Simon, 1989). Rather, the idea is to provide an overarching framework in which the host of interacting variables may be arrayed, so as to better understand how the system maintains and restores a functional equilibrium. In particular, it allows us to ask: How are the perturbations in one part of the system balanced by compensating forces in another?

For example, whereas more 'synthetic' languages, like German, rely heavily on morphological devices to convey information, others, like English, leave more to the surrounding context (Lupyan & Dale, 2010). This mode of typological inquiry can help uncover how languages use different means to nevertheless achieve similar functional ends, and the potential trade-offs—in terms of complexity and efficiency—that these different strategies may incur (Pellegrino, Coupé, & Marsico, 2011).

Two Germanic Tongues

One longstanding puzzle for typologists concerns why certain languages employ ***grammatical gender***, which assigns nouns to distinct classes and marks neighboring words for agreement. From a taxonomic standpoint, gender specification can often appear arbitrary, with little obvious correspondence between the semantic properties of a given referent and its noun class (Vigliocco et al., 2005). Historically, gender has thus been viewed as a useless ornament with little apparent rhyme or reason (Maratsos, 1979). In previous work, Dye et al. (2017) offered a possible solution to this puzzle, using an information theoretic lens to clarify the communicative function of noun classification in German.

On their account, grammatical gender marking serves to modulate nominal entropy, making nouns more equally predictable in context. This functionality benefits language processing in multiple ways: 1) by helping speakers avoid the peaks in uncertainty that would otherwise occur over nouns, smoothing entropy over the larger sequence; 2) by reducing competition between nouns that are highly confusable in context; and 3) by facilitating the use of a richer array of lexical items.

These findings raise a further puzzle: If gender systems are so beneficial to processing, languages should tend to maintain or expand them as they evolve. Yet a number of closely related Germanic tongues have followed precisely the opposite trajectory: Swedish, Danish, and Dutch have all consolidated their noun classification systems, while English has dispensed with gender altogether.

Like Modern German, Old English (~750-1150 AD) classified nouns according to three genders (masculine, feminine, and neuter) and all inanimate nouns belonged to one of the three classes (Curzan, 2003). However, in Modern English, aside from a few archaic exceptions, only nouns referring to males and females take gendered pronouns; inanimate nouns are neuter. The gender system in Modern English is thus far simpler than the noun class systems found in Old English and Modern German.

This raises the worrying possibility that English lacks the resources to accomplish the same specificity of expression available in German. However, another possibility, explored here, is that rather than employing a rigid grammatical device, English relies on a more graded, semantically transparent method of entropy reduction: namely, ***prenominal adjectives***.

Like gender markers, adjectives may act to systematically delimit the space of following nouns. For example, *massive* and *moist* are likely to have markedly different following distributions. Yet even subtle differences, such as that between *great big* and *very big*, could be highly informative in English. To test this proposal, we use tools from information theory to compare gender marking in German (a deterministic system) to prenominal adjective use in English (a probabilistic one).

Nominal Uncertainty Management

Languages appear to be organized to maintain relatively stable levels of uncertainty across discourse (Genzel & Charniak, 2002), employing various strategies to make each lexical choice more equally predictable in context, and thereby reducing processing difficulties (Tily et al., 2009; Jaeger, 2010). In information theory, uncertainty is quantified in terms of ***entropy***. Formally, the entropy H over a distribution of lexical items is a measure of the expected value of information ('surprisal') over the full range of items (Shannon, 1948):

$$H(X) = -\sum_{x \epsilon X} p(x) log_2 p(x)$$

In many languages, like English and German, nouns are the most diverse part of speech. When prior context is ignored, uncertainty should thus be highest at points where a noun occurs. For example, in the following sequence, uncertainty over possible noun continuations (*!*) will be higher than for possible verb continuations (#):

I would # like a *!* beer (1)

Unsurprisingly, nouns are among the most common sites for disfluencies, incorrect retrieval, and mishearings (Clark & Wasow, 1998; Vigliocco, 1997).

Nevertheless, speakers have various resources at their disposal for making a particular lexical choice more or less predictable in context. One possibility is to rely on the preceding discourse as a form of scaffolding. Noun class is an efficient system for implementing this principle. Consider the German equivalent of (1):

Ich hätte gern ein *!* Bier (2)

Grammatical gender markers can significantly ease the lexical access problem by systematically narrowing the set of candidate nouns that follow (Dahan et al., 2000), thereby offloading some of the uncertainty about the upcoming noun onto the determiner.

To evaluate this hypothesis, Dye et al. (2017) examined the entropy of nouns in German, a language with a three-class gender system. An analysis of the Stuttgart deWaC mega-corpus (Faaß & Eckart, 2013) revealed that gender markers systematically reduced nominal entropy across all cases. Further, this appeared to benefit ***lexical diversity***: German plurals, which are not gender-marked, showed a reduction in their type/token ratio, suggesting that the presence of a gender marker was catalyzing the use of a wider array of lexical items.

Yet English is not without its own entropy-smoothing resources. Compared to the sparse semantic context provided by (1), the noun *beer* should be more predictable following the comparatively constraining context provided by (3):

I would like a nice cold *!* beer (3)

This raises an important question: Might prenominal adjectives in English serve a similar function to grammatical gender markers in German?

Suggestive evidence comes from the visual world paradigm, an experimental framework for studying online language processing in which subjects' eye movements over a visual display are monitored as they listen to a concurrent speech stream (Tanenhaus et al., 1995). A common finding is that listeners fixate semantically-related pictures as they become relevant, with patterns of eye movements time-locked to incoming speech. In studies of this kind, prenominal adjectives and gender markers have been shown to play similar functional roles: When French and Spanish speakers encounter a gendered determiner, they rapidly shift their gaze to gender-consistent referents in the display in anticipation of the upcoming noun (Dahan et al., 2000; Lew-Williams & Fernald, 2007). Similarly, when viewing an array of semantically plausible competitors, English speakers interpret prenominal adjectives contrastively, quickly homing in on likely candidates (Sedivy et al., 1999; Fernald, Thorpe, & Marchman, 2010). Such findings suggest that both prenominal elements serve a predictive, discriminative function.

Corpus Analysis

To more closely examine this apparent functional similarity between languages, we conducted a comparison of prenominal adjective and determiner usage in written English and German.

Corpora

Analyses were initially run on manually annotated newswire corpora, and subsequently replicated on larger web-crawled mega-corpora. These corpus types trade off on scale and precision. Due to space constraints, we report one or the other, but not both; in each case, the qualitative nature of the results are the same.

The newswire corpora included the Negra II corpus of German newspapers, (Skut et al. 1997) and the New York Times Gigaword corpus (Graff et al., 2007). The web-crawled WaCky mega-corpus supplied the SdeWaC, a subset of the German section (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009), comprising more than 850M word tokens and 1.1 M word types (Faaß & Eckart, 2013), and the ukWaC, the British English subset, comprising nearly 2 billion word tokens and 3.8 M word types (Ferraresi et al., 2008). It is worth noting that these are collections of written language, which may not reflect the complexities of spoken production (Baayen, Milin, & Ramscar, 2015).

Additional annotation for fine-grained part-of-speech categories and extraction was carried out with the RFTagger (Schmid & Laws, 2008) and the Stanford Parser (Klein & Manning, 2003).

Determiners

Entropy Reduction In German, grammatical gender serves to subdivide the space of nouns that can legally follow each marker. By markedly reducing nominal entropy, gender facilitates the use of a more diverse—and more informative—set of nouns following gender-marked determiners. Consistent with this thesis, when Dye et al. (2017) compared singular nouns in German (which are marked for gender), with plural nouns (which are not), they found that singular nouns following determiners were significantly more lexically diverse than their plural counterparts.

By comparison, English determiners, which are neither gender nor case-specific, have less potential to be informative about their following nouns. Consider that while the determiner *the* in English is informative about the type of word that will follow (a noun, most likely), in German, the determiners *der*, *die*, *das, den, dem*, and *des* convey not only part of speech information, but also delineate the specific set of lexical items that can follow. This suggests that English determiners may not support the same level of lexical diversity available in German.

To examine this possibility, we first compared the conditional entropy of German nouns following articles (which are gender-marked) to that of English nouns following articles (which are not), in the Negra II and NYT Gigaword corpus, respectively. While the average uncertainty following the determiners was similar across languages, German determiners supported much greater entropy reduction than their English equivalent, a result that held across corpus types.

As Figure 1 illustrates, following a definite article, the conditional entropy of English nouns was similar to that of German nouns (10.17 vs. 10.55). However, whereas German provided a substantial entropy offset, English provided none at all. In German, removing information about definite articles—and hence, about noun class—led to a significant increase in entropy (from 10.55 to 11.71 bits). In the simplified model corpus depicted in Figure 1, whereas the baseline entropy difference between marked English and German nouns suggests a usage rate of around 30% more nouns, the difference between marked and unmarked German nouns is the equivalent of more than 125% more nouns.

Lexical Diversity This finding suggests that compared to English, German noun usage must be more heterogeneous following determiners (**Figure 1**). To compare nominal usage across languages, we calculated the type/token ratio of noun lemmas in these contexts in the Negra II and NYT Gigaword corpus, following Dye et al. (2017), and normalizing for corpus size to make the results comparable. Conveniently, type/token ratio is the inverse of average frequency, which means that the greater the diversity of nominal usage, the lower the average frequency. We found that whereas the average frequency of the German noun lemmas in Negra II was 2.12, the average frequency of similar noun lemmas in the English ukWaC sample was 4.93 ($p < 0.001$).

These results suggest that noun class allows German speakers to use more 'informative'—and therefore, less frequent and less predictable—nouns after definite articles more often than English speakers do. Or, to put it another way, German speakers appear to use the entropy reduction provided by noun class to choose nouns that are more specific, resulting in greater nominal diversity.

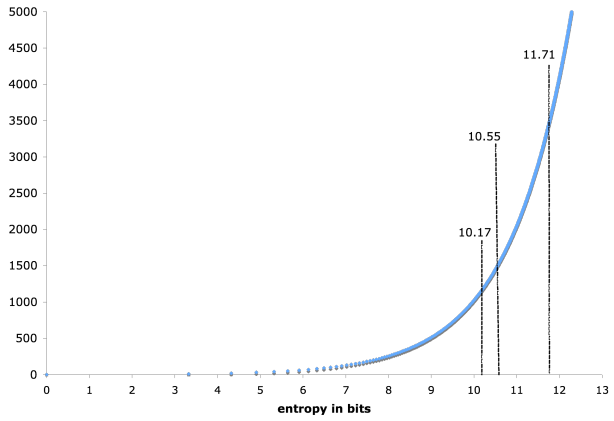

Figure 1: To illustrate the relationship between entropy, probability, and frequency in a corpus of nouns, the x-axis above represents the entropy for a given noun as the size of a set of nouns of equal frequency (1, y-axis) increases. As the size of the set of items increases linearly, entropy rises as an exponential function.

Adjectives

While our results confirm that nominal usage following determiners is more diverse in German than in English, it does not therefore follow that English is lexically impoverished compared to German, or unable to achieve the same degree of specificity. In particular, definite articles are not the only type of word that typically precede nouns—

adjectives are also common prenominally, and may serve a similar function.

To further explore this idea, we compared the adjective-noun sequences in the ukWaC and the SdeWaC corpora. Both the overall proportion of adjectives (t_p = 1992.336; p < .0001) and the probability of a noun being preceded by an adjective (t_p = 85.088; p < .0001) were significantly higher in English than in German. While German nouns are significantly more lexically diverse than their English counterparts, precisely the opposite obtains for adjectives.

These results present us with two different theoretical explanations. One possibility is that English speakers use more adjectives overall to compensate for the use of less varied nouns in communication—i.e., that they make their messages more specific through adjectival syntagmatic choices. Alternatively, it might be that while German speakers use more varied nouns after articles, English speakers use an equally diverse set of nouns, but rely on adjectives—rather than determiners—to facilitate the use of more informative nouns. Fortunately, these accounts make competing predictions, allowing us to distinguish between them empirically.

The first account accords well with the taxonomic assumption that adjectives add semantic detail to nouns, or somehow "modify" their semantic content (Kamp & Partee, 1995). On this assumption, adjectives should preferentially modify high frequency nouns, which are in greater need of semantic augmentation, over low frequency nouns, which tend to be more specific (Rosch, 1978). For example, *dog* is less informative than *retriever*, which is less informative than *dachshund*; accordingly, *dog* should be the most frequently modified, and *dachshund* the least.

However, if prenominal adjectives in English serve a similar role to gendered determiners in German, precisely the opposite prediction should be made regarding frequency. In German, the entropy reduction properties afforded by noun class facilitate the use of more informative (lower frequency) nouns. If, in English, at least some of this functionality is subsumed by prenominal adjectives, then it is *low* frequency nouns that should be preferentially "modified", not high frequency ones. The relationship between adjectives and noun frequency thus provides an important test case.

In line with the entropy smoothing account, our analysis reveals a negative correlation between a noun's log frequency and its likelihood of being modified (r = -0.17, p < 0.001). Moreover, our investigation[1] indicates that in English, adjectives redistribute the relative entropy of nouns, thus serving to balance the degree to which nouns can be predicted in context: More frequent nouns tend to be preceded by adjectives that are (on average) higher frequency and higher entropy (*edf* = 22.06: F = 32069: p < 0.0001). Indeed, a nonlinear interaction between adjective entropy and adjective frequency accounts for fully 94% of the variance in noun frequency (**Figures 3, 4**).

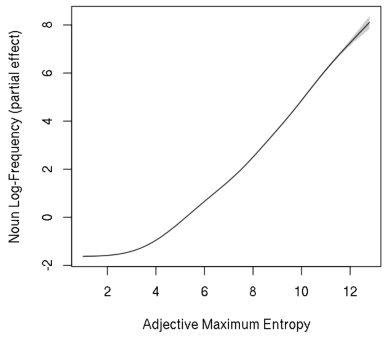

Figure 3: Adjective maximum entropy, which provides an upper bound on uncertainty about the preceding adjective, accounted for almost 90% of the variance in noun frequency (*Adjusted* R^2 = 0.891), revealing that more frequent nouns are preceded by larger number of different adjectives.

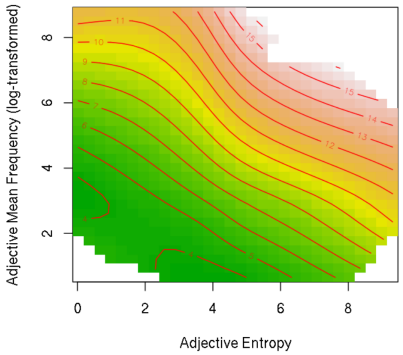

Figure 4: To achieve greater precision, a second interactive model was run, which regressed noun frequency with the tensor product of adjective entropy by adjective average frequency. This interactive model accounted for fully 94% of the variance in noun frequency (*Adjusted* R^2 = 0.941), and achieved better goodness-of-fit than the max entropy model, as indicated by both the difference in *AIC* (28603.60), and the Chi-square test of fREML scores (χ^2 = 14274.832, *edf difference* = 3.000, p < 0.0001).

Discussion

In comparing English and German, two closely related Germanic tongues, we found that whereas German nouns are significantly more lexically diverse than their English counterparts, precisely the opposite obtains for adjectives. These results suggest that the difference between German and English does not lie in the 'specificity' of expression, per se, but rather in how specificity is achieved.

German uses gender marking to distinguish between likely lexical competitors, and adjectives to make rarer lexical items more predictable in context. By contrast, in English, which largely lacks gender, adjectives assume both roles. While these findings are compatible with discriminative accounts of language processing (Ramscar et al, 2010), they raise questions about the explanatory adequacy of traditional taxonomic theories.

[1] To better understand adjective-noun relations in English, we retrieved the top 50k most frequency nouns from the ukWaC, along with information about the adjectives preceding them, including: their number, average frequency, and entropy (i.e., the uncertainty over the noun's prior distribution). These results were log-transformed to approximate normality. For convenience, a base-two logarithmic transform was used for adjective-number, yielding adjective maximum entropy in bits.

General Discussion

To Gender or Not

On an evolutionary scale, languages tend to become more codified over time, as frequently used sequences of words gradually crystallize into more rigid conventions, a process known as grammaticalization (Hopper & Traugott, 1993). However, at a number of points in its history, English has taken the opposite developmental path.

One such turning point was the invasion and colonization of the British Isles in the 8th and 9th centuries by the Norse, followed by the Norman conquest of England in the 11th century. As a result of the extended interaction between Old English and Norse, much of the information that had been encoded in fixed aspects of the grammar became "optional" – expressed by words rather than fixed grammatical markers. Old English, the language of England at the beginning of this period, looks like Modern German, with relatively complex patterns of inflection for number, gender, and case. However, by the end of this period, Old English had been eclipsed by Middle English, which much more closely resembles the modern tongue: nouns are marked only for number, adjectives are no longer inflected, and demonstratives are reduced in kind (Dawson, 2003). What might explain this trajectory?

Evolutionary Pressures While it is well known that some languages are easier for adult learners to master than others, it is also the case that first languages are acquired at different rates—Russian children, for example, take several years longer than their Turkish neighbors to sort out nominal case marking (Slobin, 2006). However, what is difficult for a child to learn, may not be difficult for an adult, and vice versa; early language acquisition and adult second-language learning are qualitatively different, both in the nature of the task demands, and in the capacities of the learners themselves (Ramscar & Gitcho, 2007; Thompson-Schill et al., 2009). Likewise, there may be tradeoffs between what is easy to acquire, and what is efficient to process (see Ramscar et al. 2010 on adjective ordering).

In line with this proposal, there is accruing evidence that the structural form of a language is coupled to its population (and history) of adult learners (Johnson & Newport, 1989; Trudgill, 2002). Support for this comes from a series of in-depth analyses of the World Atlas of Language Structures conducted by Lupyan and Dale (2010), who found that languages with "larger speaker populations, greater geographical coverage, and greater degree of contact with other languages" (p. 6) tend to be morphologically simpler, more transparent in their mappings between form and meaning, and more likely to express semantic distinctions through lexical or pragmatic means, rather than encoding them explicitly in the grammar.

On this account, languages strike a balance between early learnability and adult processing that is moderated by their social niche. Thus, while morphologically complex languages provide a rich set of additional cues to scaffold infant learning, this early advantage has significant drawbacks for adult speakers. The same marking conventions that support young learners, prove nearly impossible for adult learners to master (Johnson & Newport, 1989), particularly when extrapolating from noisy input (Hudson Kam & Newport, 2009).

In languages spoken by large populations of adult learners, there is thus both impetus and imperative to simplify the obligatory aspects of the grammar. Moreover, adult speakers are instrumental to how languages evolve—it is skilled language users (not novices) who make and spread innovations (Labov, 1972; Trudgill, 2010), and adult learners readily adapt newly acquired grammars to better meet their communicative needs (Fedzechkina, Jaeger, & Newport, 2011; Kirby, Cornish, & Smith, 2010).

From this perspective, the distribution of Modern English can be seen as having developed in response to the selective pressures produced by the conflicting gender systems of Old English and Norse, combined with a large percentage of adults in the population of language learners. These conditions resulted in a shift away from the abstract, grammaticalized entropy management system of Old English gender marking, to the more probabilistic, semantically transparent system based on adjectives found in Modern English. In comparison to German and Old English, Modern English has thus traded efficiency—in communicative terms—for error tolerance, making it more amenable to later learning.

Adjectives and Overspecification

From a certain perspective, languages with complex inflectional patterns can appear inefficient, in that they obligate the marking of certain distinctions—such as the temporal remoteness of an action or event—that may or may not actually be relevant to the topic at hand (Lupyan & Dale, 2010). Yet languages with more transparent semantics employ much of the same apparent redundancy: Native English speakers, who are not grammatically obliged to be superfluous, still regularly produce overspecified utterances like "that's a *cute little* puppy" and "how about a *nice cold* beer?" (Deutsch & Pechmann, 1982; Engelhardt et al., 2006) The logic of such productions has proved notoriously difficult to account for: For one, they appear to violate the Gricean Maxim of Quantity, which assumes that speakers provide just enough information to identify a referent, and no more; for another, their combinatorial meaning has defied systematic description (Lahav, 1989), relegated by formal semanticists to the realms of 'context dependence' and 'vagueness' (Kamp & Partee, 1995).

However, productions like these only appear mysterious if their meanings are assumed to be compositional—i.e., constructed as a function of their syntax and the meanings of their constituent parts (Fodor & Lepore, 2002; but see Baroni & Zamparelli, 2010 for a novel approach). Under the alternative model suggested by information theory, utterances are produced so as to iteratively reduce uncertainty, and different languages employ more (or less) conventionalized means of streamlining that process (Baayen & Ramscar, 2015). While the patterns of adjective use in English are difficult to account for in terms of formal semantics, their communicative function is strikingly clear from an information theoretic perspective.

Future Directions

One straightforward extension of this work is to the ordering of prenominal adjectives in English. Violations of conventional adjective ordering can make interpretation difficult, as when we compare 'old French red wine crates' with 'red French old wine crates.' Yet adjective ordering cannot be explained by a simple syntactic rule, and while various elaborate semantic hierarchies have been suggested (**Table 1**), they are not consistent enough to be implemented computationally (Malouf, 2000).

Modification zones for adjectives				
Determiner	Specification	Description	Classification	Head noun
The	same	beautiful	italian	actress
Her	own	handsome	jazz	musician
The	next	big	guardian	dog
This	particular	tall	aboriginal	carving
The	same	gray	church	tower

Table 1: A semantic account of adjective order, in which specifying adjectives "single out or quantify the referent", descriptive adjectives "characterize the referent along a variety of semantic parameters", and classifying adjectives "categorize the referent" (Kemmerer et al., 2007: 240).

Ziff (1960) proposed that adjective order was determined according to two closely related heuristics: the adjective's "privilege of occurrence" (i.e., the range of nouns it might modify) and its "definiteness of denotation" (i.e., the extent to which its interpretation depended on the noun being modified). On this account, adjectives that are more privileged and more definite should be slotted closer to the noun. In a related vein, Danks and Glucksberg (1971) argued that adjectives are ordered according to their "discriminative potential", with the most broadly discriminating being placed first. Both of these claims are amenable to further scrutiny in terms of information and prediction.

One possibility is that adjective chains follow the familiar branching structure seen in personal names (Ramscar et al., 2014), with set-size increasing as a function of proximity to the head noun. This would be consistent with the finding that more frequent adjectives tend to precede less frequent ones. However, given that adjectives' appear to smooth entropy, rather than simply reduce it, the precise chaining structure may be closely tied up with the frequency of the noun being 'modified'. This could explain apparent exceptions to this trend (like "witty young lawyer").

In addition to adjective order, similar analyses might help explain the cross-linguistic differences that have been observed in languages with postnominal adjective biases (Percy et al. 2009; see also Lambert & Paivio, 1956). More ambitious extensions could be made to other parts of speech, such as verbs and adverbs, and for other languages, beyond those studied here.

Acknowledgments

Many thanks are due to Christian Adam for his heroic feats of data collection in mining the WaCky corpora.

References

Arnon, I. & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order of acquisition affects what gets learned. *Cognition*, 122(3), 292-305.

Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections--phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, 54(4), 513–526.

Baayen, R. H., & Ramscar, M. (2015). Abstraction, storage and naive discriminative learning. In Dabrowska, E., and Divjak, D. (Eds.) *Handbook of Cognitive Linguistics*., 99-120. Berlin: De Gruyter Mouton.

Baayen, R. H., Milin, P., & Ramscar, M. (2015). Frequency in lexical processing. *Aphasiology*, 30(11), 1174–1220.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta. E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3), 209- 226.

Baroni, M., & Zamparelli, R. (2010). Nouns are vectors, adjectives are matrices: representing adjective-noun constructions in semantic space. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1183–1193.

Beckner, C. et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(1), 1-26.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(05), 489–509.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201-242.

Curzan, A. (2003). *Gender shifts in the history of English*. Cambridge: Cambridge University Press.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic Gender and Spoken-Word Recognition in French. Journal of Memory and Language, 42(4), 465–480.

Dahl, Ö. (2004). Studies in Language Companion Series : Growth and Maintenance of Linguistic Complexity. John Benjamins Publishing Company.

Danks, J. H., & Glucksberg, S. (1971). Psychological scaling of adjective orders. *Journal of Verbal Learning and Verbal Behavior*, 10(1), 63–67.

Dawson, H. C. (2003): Defining the Outcome of Language Contact: Old English and Old Norse. *OSUWPL*, 57, 40-57.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2017). A functional theory of gender paradigms. In F. Kiefer, J.P. Blevins, & H. Bartos (Eds.) *Perspectives on Morphological Organization: Data and Analyses*. Brill: Leiden.

Engelhardt, P., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554–573.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(05), 429–448.

Faaß, G., & Eckart, K. (2013). SdeWaC - A Corpus of Parsable Sentences from the Web. Gurevych, Iryna, Chris Biemann & Torsten Zesch (eds.): *GSCL 2013, LNCS 8105*. (Heidelberg: Springer).

Fedzechkina, M., Jaeger, T.F., & Newport, E.L. (2011). Functional biases in language learning: Evidence from word order and case-marking interaction. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX.

Fernald, A., Thorpe, K., & Marchman, V.A. (2010). Blue Car, Red Car: Developing Efficiency in Online Interpretation of Adjective-Noun Phrases. *Cognitive Psychology*, 60(3), 190-217.

Ferraresi, A., Zanchetta, E., Baroni, M. & Bernardini, S. (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop* (WAC-4). 47–54.

Fodor, J.A. & Lepore, E. (2002). *The Compositionality Papers*. Oxford University Press.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text (pp. 199–206). In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics: Morristown, NJ.

Graff, D. et al. (2007). English Gigaword Third Edition LDC2007T07. Web Download. Philadelphia: Linguistic Data Consortium.

Hopper, P., & Traugott, E. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62.

Kamp, H. & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57, 129-191.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory. *PNAS*, 105(31), 10681–10686.

Kemmerer, D., Weber-Fox, C., Price, K., Zdanczyk, C., & Way, H. (2007). Big brown dog or brown big dog? An electrophysiological study of semantic constraints on prenominal adjective order. *Brain and Language*, 100(3), 238–256.

Labov, S. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Lahav, R. (1989). Against compositionality: The case of adjectives. *Philosophical Studies*, 57(3), 261–279.

Lambert, W. E., & Paivio, A. (1956). The influence of noun-adjective order on learning. *Canadian Journal of Psychology*, 10(1), 9–12.

Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS ONE*, 5(1), e8559.

Malouf, R. (2000). The order of prenominal adjectives in natural language generation. *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 85-92.

Maratsos, M. P. (1979). How to get from words to sentences. In D. Aaronson & R. Rieber (eds.), *Perspectives in psycholinguistics*. Hillsdale, N.J.: Erlbaum

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140(3), 325–347.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27(1).

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3), 539–558.

Percy, E. J., Sherman, S. J., Garcia-Marques, L., Mata, A., & Garcia-Marques, T. (2009). Cognition and native-language grammar: The organizational role of adjective-noun word order in information representation. *Psychonomic Bulletin & Review*, 16(6), 1037–1042.

Ramscar, M. & Baayen, H. (2013). Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Sciences*, 4, 233.

Ramscar, M. & Gitcho, N. (2007) Developmental change and the nature of learning in childhood. *Trends In Cognitive Science*, 11(7), 274-279.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34, 909-957.

Ramscar, M., Smith, A.H., Dye, M., Futrell, R., Hendrix, P., Baayen, H. & Starr, R. (2013). The 'universal' structure of name grammars and the impact of social engineering on the evolution of natural information systems. *Proceedings of the 35th Meeting of the Cognitive Science Society*, Berlin, Germany.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillside, NJ: Lawrence Erlbaum Publishers.

Schmid, H. & Laws, F. (2008). Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-Grained POS Tagging. COLING 2008: Manchester, Great Britain.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423, 623–656.

Simon, H. A. (1989). Cognitive Architectures and Rational Analysis: Comment. *Technical Report AIP*, 58, 1–25.

Skut, W., B. Krenn, T. Brants., & H. Uszkoreit. (1997). An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing* (ANLP).

Slobin, D. (2006). Cross-linguistic comparative approaches to language acquisition. *Encyclopedia of Language & Linguistics*, 299–301.

Tanenhaus M. K., Spivey-Knowlton M. J., Eberhard K. M., Sedivy J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science 268 1632–1634.

Thompson-Schill, S., Ramscar, M., & Chrysikou, E. (2009) Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, 8(5), 259-263.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(2), 147–165.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Trudgill, P. (2002). *Sociolinguistic Variation and Change*. Edinburgh University Press.

Trudgill, P. (2010). *Investigations in sociohistorical linguistics: Stories of colonisation and contact*. Cambridge University Press.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical Gender Is on the Tip of Italian Tongues. *Psychological Science*, 8(4), 314–317.

Vigliocco, G., Vinson, D.P., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology: General*, 134, 501-520.

Ziff, P. (1960). *Semantic Analysis*. Cornell University Press: Ithaca, NY.